(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 065 824 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
03.06.2009 Bulletin 2009/23

(51) Int Cl.:
G06F 21/00 *(2006.01)*

(21) Application number: **08020827.5**

(22) Date of filing: **01.12.2008**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR
Designated Extension States:
AL BA MK RS

(30) Priority: **30.11.2007 US 991504 P**

(71) Applicant: **Fox Entertainment Group**
Los Angeles, CA 90067 (US)

(72) Inventors:
- **Kunal, Anand**
  Los Angeles, CA 90067 (US)
- **Kaminsky, Dan**
  Los Angeles, CA 90067 (US)

(74) Representative: **Delorme, Nicolas et al**
Cabinet Germain & Maureau
BP 6153
69466 Lyon Cedex 06 (FR)

(54) HTML filter for prevention of cross site scripting attacks

(57) An HTML filter is described that converts HTML tags into HTML object and associated param tags. In an exemplary embodiment, the present HTML filter also validates existing object tags so that they may render in at least one, and optionally all, major browsers. In another exemplary embodiment, the presently described HTML filter also serves as a configurable whitelist for rich media (through controlling particular attributes, e.g., "classid", in the object tag and affiliated param tags).

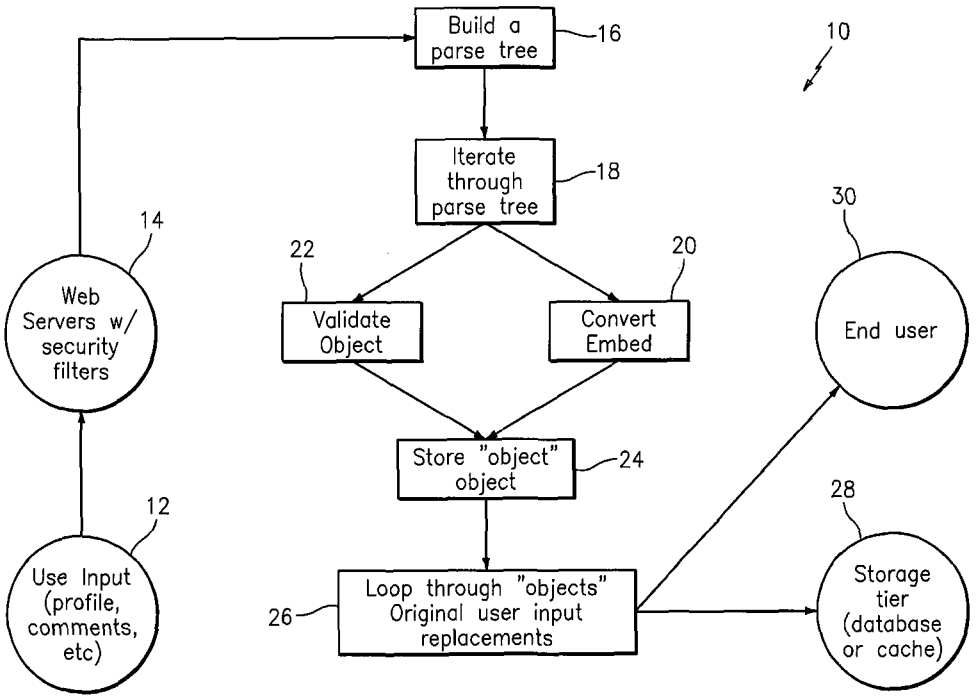

FIG. 1

EP 2 065 824 A1

Description

BACKGROUND

[0001] Internet web sites, and in particular, social networks, have evolved into media rich experiences. While allowing users to embed rich media creates a more engaging online environment, such embedded rich media presents certain security risks. Specifically, such embedding of rich media forces website administration to sanitize content to prevent cross site scripting (XSS) attacks that might otherwise occur. However, given the amount of user-generated input on such sites as well as the server's ability to manipulate headers for file extensions, it is difficult if not impossible to crawl and check the validity of remote data (headers and file analysis).

[0002] What is needed in the art are effective mechanisms for preventing such cross site scripting attacks without neglecting to address embed and object tags.

SUMMARY

[0003] The present invention recognizes that embed tags are inherently insecure and can allow remote code execution on a web site through various file formats, including but not limited to Quicktime, Adobe PDF, etc.

[0004] Similarly, the present invention recognizes that, provided exact construction of both object and param tags, object tags will not only render rich media, but will not execute remote code.

[0005] Accordingly, the presently described HTML filter converts HTML tags into HTML object and associated param tags. In an exemplary embodiment, the present HTML, filter also validates existing object tags so that they may render in at least one, and optionally all, major browsers. In another exemplary embodiment, the presently described HTML filter also serves as a configurable whitelist for rich media (through controlling particular attributes, e.g., "classid", in the object tag and affiliated param tags).

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Referring now to the drawings, wherein like elements are numbered alike in the following FIGURE:

[0007] FIGURE 1 is an exemplary flowchart illustrating an exemplary HTML filter process for conversion of embedded rich media to object tags and associated param tags.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0008] Reference will now be made in detail to exemplary embodiments, examples of which are illustrated by the accompanying drawing.

[0009] As is noted above, the presently described HTML filter converts HTML tags into HTML object and associated param tags. In an exemplary embodiment, the present HTML filter also validates existing object tags so that they may render in at least one, and optionally all, major browsers. In another exemplary embodiment, the presently described HTML filter also serves as a configurable whitelist for rich media (through controlling particular attributes, e.g., "classid", in the object tag and affiliated param tags).

[0010] Referring now to FIGURE 1, the illustrated flowchart shows an exemplary HTML filter process 10 for conversion of embedded rich media to object tags and associated param tags. In a first step, a user inputs rich media, profile data, comments, etc. at 12. At step 14, the web server runs one or more security filters with regard to the imputed data. The server then builds a parse tree 16 and iterates through the parse tree 18. The server then converts embeds into objects 20, validates objects 22 and stores objects converted from embeds 24. In a further step, the server loops through the stored objects and provides the replacements for the original user input 26. The server then stores the modified input 28 and ends the process 30.

[0011] Any of the above described HTML filter configurations will prevent remote code execution or worms with regard to rich media embeds. This will result in significant cost reductions (i.e., time spent for investigations, data cleanup, monetary damages, etc.). Additionally, the above described HTML filter alleviates the need to automatically block typically risky rich media types (e.g., Apple's QuickTime is a particularly target rich engine for site attacks, and it is often entirely blocked). Thus, the system may be configured, with the above described HTML security filter, to accept such (and process) such media without risk of cross site scripting attacks.

[0012] It will be apparent to those skilled in the art that, while exemplary embodiments have been shown and described, various modifications and variations can be made to the HTML filter for prevention of cross site scripting attacks as is disclosed herein without departing from the spirit or scope of the invention. Accordingly, it is to be understood that the various embodiments have been described by way of illustration and not limitation.

Claims

1. A method for modifying hypertext markup language to prevent cross site scripting attacks, comprising:

   providing a filter for said hypertext markup language, wherein said filter acts on hypertext markup language tags and converts said tags into hypertext markup language object tags and associated parameter tags to prevent cross site scripting attacks.

2. A method in accordance with claim 1, wherein said hypertext markup language filter also validates ex-

isting object tags so that they are configured to render in a plurality of browsers.

3. A method in accordance with claim 1, wherein said hypertext markup language filter further acts as a configurable whitelist for rich media.

4. A method in accordance with claim 3, wherein said hypertext markup language filter controls particular attributes in the object tag and affiliated parameter tags to provide said whitelist.

5. A method in accordance with claim 4, wherein said hypertext markup language filter controls "classid" attributes in the object tag and affiliated parameter tags.

6. A method in accordance with claim 2, wherein said hypertext markup language filter acts as a configurable whitelist for rich media.

7. A method in accordance with claim 6, wherein said hypertext markup language filter controls particular attributes in the object tag and affiliated parameter tags to provide said whitelist.

8. A method in accordance with claim 7, wherein said hypertext markup language filter controls "classid" attributes in the object tag and affiliated parameter tags.

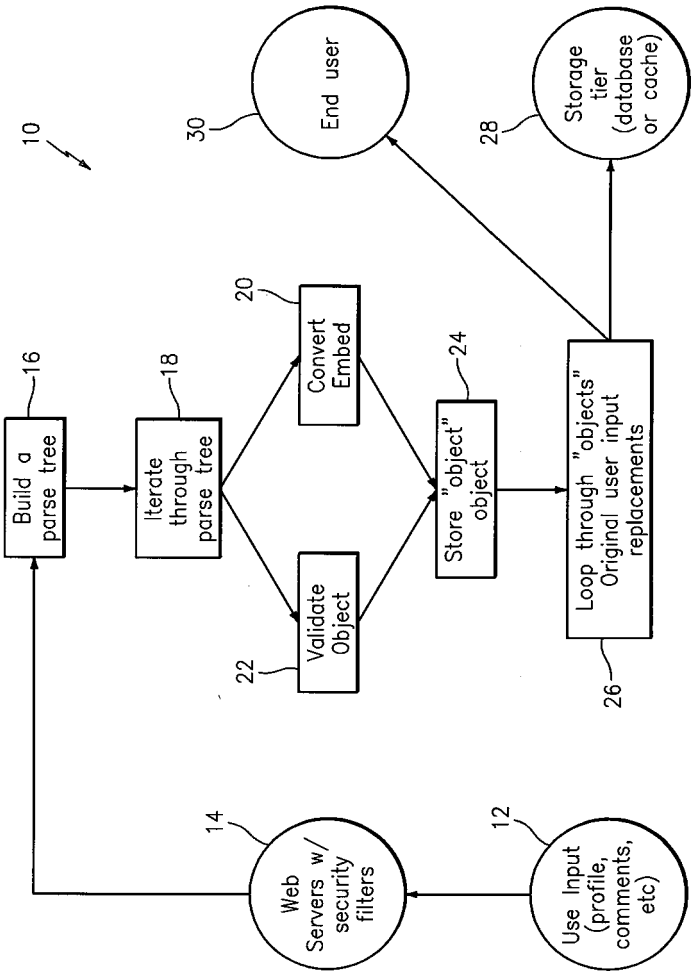
10
30
End user
28
Storage tier (database or cache)
20
Convert Embed
16
18
Build a parse tree
Iterate through parse tree
24
Store "object" object
Loop through "objects" Original user input replacements
22
Validate Object
26
14
Web Servers w/ security filters
12
Use Input (profile, comments, etc)

FIG. 1

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 08 02 0827

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2005/198692 A1 (ZURKO MARY E [US] ET AL) 8 September 2005 (2005-09-08) * abstract * * figures 1-4 * * paragraphs [0003], [0004] * * paragraphs [0016], [0017] * * paragraphs [0020] - [0031] *	1-8	INV. G06F21/00
X	OBSCURE: "Bypassing JavaScript Filters - the Flash! Attack"[Online] 25 August 2002 (2002-08-25), XP002521381 Retrieved from the Internet: URL:http://eyeonsecurity.org/papers/flash-xss.htm> [retrieved on 2009-03-26] Section Vulnerable site and software examples	1-8	
X	EUROSEC GMBH: "Filtering JavaScript to Prevent Cross-Site Scripting"[Online] 1 July 2005 (2005-07-01), XP007907964 Retrieved from the Internet: URL:http://www.secologic.org/downloads/web/051207b_EUROSEC_Draft_Whitepaper_Filtering_JavaScript.pdf> [retrieved on 2009-03-27] Section 2.2 Section 3	1-8	TECHNICAL FIELDS SEARCHED (IPC) G06F

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	30 March 2009	Schäfer, Andreas

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 08 02 0827

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	S. YOSHIHAMA, A. ISHIDA, N. URAMOTO: "Beyond XSS - Towards Universal Content Filtering" IBM RESEARCH REPORT, [Online] 2 November 2007 (2007-11-02), XP002521383 Retrieved from the Internet: URL:http://domino.research.ibm.com/library/cyberdig.nsf/papers/B1FC30F8EE161B4D8525739A001A13CC/$File/RT0764.pdf> [retrieved on 2009-03-26] * pages 1-3 * * abstract * -----	1-8	
A	JAYAMSAKTHI SHANMUGAM ET AL: "XSS Application Worms: New Internet Infestation and Optimized Protective Measures" SOFTWARE ENGINEERING, ARTIFICIAL INTELLIGENCE, NETWORKING, AND PARALLEL/DISTRIBUTED COMPUTING, 2007. SNPD 2007. EIGHTH ACIS INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 1 July 2007 (2007-07-01), pages 1164-1169, XP031125251 ISBN: 978-0-7695-2909-7 Section 4 Section 5 * pages 1-4 * -----	3,4,6,7	TECHNICAL FIELDS SEARCHED (IPC)
A	US 2007/107057 A1 (CHANDER AJAY [US] ET AL) 10 May 2007 (2007-05-10) * figures 8-11 * * paragraph [0051] * -----	1-8	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	30 March 2009	Schäfer, Andreas

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 08 02 0827

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

30-03-2009

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005198692 A1	08-09-2005	NONE	
US 2007107057 A1	10-05-2007	EP 1955249 A2	13-08-2008
		EP 1962220 A1	27-08-2008
		WO 2007058882 A2	24-05-2007

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82